Four-class ASME BCI: investigation of the feasibility and comparison of two strategies for multiclassing

Simon Kojima[1]* and Shin'ichiro Kanoh[1,2]

[1]Graduate School of Engineering and Science, Shibaura Institute of Technology, Tokyo, Japan,
[2]College of Engineering, Shibaura Institute of Technology, Tokyo, Japan

Introduction: The ASME (stands for Auditory Stream segregation Multiclass ERP) paradigm is proposed and used for an auditory brain-computer interface (BCI). In this paradigm, a sequence of sounds that are perceived as multiple auditory streams are presented simultaneously, and each stream is an oddball sequence. The users are requested to focus selectively on deviant stimuli in one of the streams, and the target of the user attention is detected by decoding event-related potentials (ERPs). To achieve multiclass ASME BCI, the number of streams must be increased. However, increasing the number of streams is not easy because of a person's limited audible frequency range. One method to achieve multiclass ASME with a limited number of streams is to increase the target stimuli in a single stream.

Methods: Two approaches for the ASME paradigm, ASME-4stream (four streams with a single target stimulus in each stream) and ASME-2stream (two streams with two target stimuli in each stream) were investigated. Fifteen healthy subjects with no neurological disorders participated in this study. An electroencephalogram was acquired, and ERPs were analyzed. The binary classification and BCI simulation (detecting the target class of the trial out of four) were conducted with the help of linear discriminant analysis, and its performance was evaluated offline. Its usability and workload were also evaluated using a questionnaire.

Results: Discriminative ERPs were elicited in both paradigms. The average accuracies of the BCI simulations were 0.83 (ASME-4stream) and 0.86 (ASME-2stream). In the ASME-2stream paradigm, the latency and the amplitude of P300 were shorter and larger, the average binary classification accuracy was higher, and the average weighted workload was smaller.

Discussion: Both four-class ASME paradigms achieved a sufficiently high accuracy (over 80%). The shorter latency and larger amplitude of P300 and the smaller workload indicated that subjects could perform the task confidently and had high usability in ASME-2stream compared to ASME-4stream paradigm. A paradigm with multiple target stimuli in a single stream could create a multiclass ASME BCI with limited streams while maintaining task difficulty. These findings expand the potential for an ASME BCI multiclass extension, offering practical auditory BCI choices for users.

1 Introduction

The brain-computer interface (BCI) allows users to control external devices without muscle activation by decoding their intention from neural signals. Numerous studies have been conducted on BCI for patients who have severe motor impairments, completely locked-in syndrome (CLIS), amyotrophic lateral sclerosis (ALS) or spinal cord injury (SCI) (Wolpaw et al., 2000; King et al., 2013; Guger et al., 2017; Zhang et al., 2017; Guger et al., 2020) and for healthy people (Holzner et al., 2009; Kosmyna et al., 2016; Yu et al., 2016; Pan et al., 2017; Gao et al., 2018; Park et al., 2019). Electroencephalogram (EEG) is widely used to measure brain activity because they are suitable for measuring signals with high temporal resolution via portable devices (Vidal, 1973; Wolpaw et al., 2002; Rao, 2019).

Among EEG-based BCIs, synchronous BCIs detect a stereotypical brain response generated after the subject is presented with a stimulus (Rao, 2019). Some synchronous BCIs that use visual stimuli have been proposed. These visual BCIs have performed well (Cheng et al., 2002; Gao et al., 2003; Bin et al., 2009; Thielen et al., 2021; Martínez-Cagigal et al., 2021). However, BCIs based on visual stimuli occupy the user's sight, and thus, cannot be used by patients with visual impairments.

In contrast, auditory BCIs, which use auditory stimuli, do not require visual modality and can be used without restricting the user's visual functions. Hill et al. proposed a famous example of an early auditory BCI (Hill et al., 2004). They presented two different oddball sequences to the right and left ears of the subject. In an oddball paradigm, subjects are presented with a sequence of frequent tones (i.e., standard or nontarget stimuli). In contrast, some are infrequently replaced with different stimuli (i.e., deviant or target stimuli). When a listener focuses on a sequence (e.g., by counting the number of target stimuli), event-related potentials, such as P300 and N200, are elicited by the target stimulus presented (Luck, 2014). In their system, the direction of the subject's focus was estimated by detecting ERP responses to the target stimuli using a support vector machine (SVM). Furdea et al. (2009) proposed the auditory speller BCI. In this system, a five-by-five matrix consisting of alphabet letters was coded with acoustically presented numbers. The subjects selected the target character by focusing selectively on the sound stimuli corresponding to the row and column numbers in sequence. This system was also tested on locked-in patients (Kübler et al., 2009). Another system was proposed by Schreuder et al. (2010). The subjects were surrounded by eight loudspeakers at the ear level, and the target direction of the user's attention was detected. Musso et al. (2022) proposed a BCI-based language rehabilitation system for poststroke aphasia patients with this paradigm. Höhne et al. (2011) proposed a 9-class auditory BCI by presenting sound stimuli with three pitches from three directions. The target direction and pitch were detected using Fisher discriminant analysis (FDA).
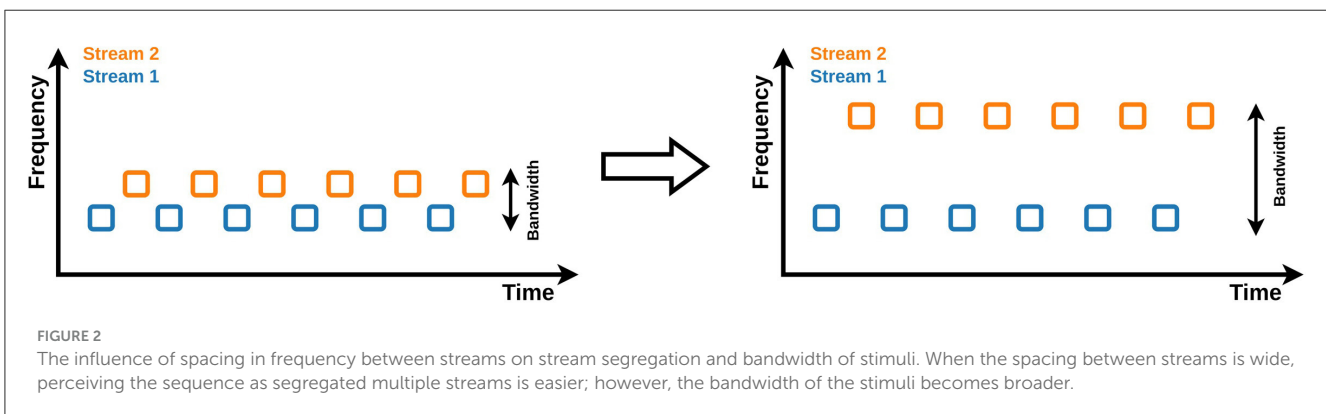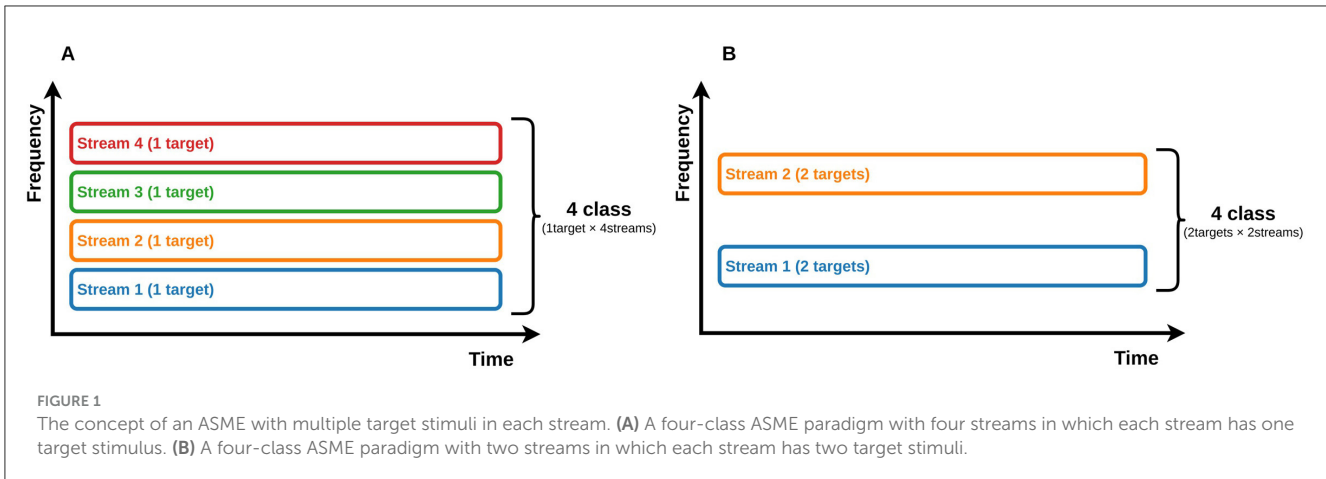
Despite many proposed auditory BCIs, their performance is lower than that of visual-based BCIs (Furdea et al., 2009; Belitski et al., 2011; Oralhan, 2019). However, patients with late-stage ALS are known to have unreliable gaze control (Choi et al., 2023); thus, proposing systems that are not dependent on visual modalities is crucial for these patients.

Furthermore, this approach may not be the most suitable choice for some users since it occupies user vision, which is a sensory modality frequently used in daily life. Therefore, developing a high-performance and practical auditory BCI is important.

The auditory BCIs mentioned above use relatively simple paradigms, such as distinguishing between high- and low-pitched tones or selecting one sound from spatially located sounds. However, the human auditory system is capable of very complex processing and can highly discriminate between various sounds, and high-performance auditory BCIs may be achieved if we can fully utilize human auditory abilities. Thus, an auditory BCI based on stream segregation was proposed (Kanoh et al., 2008, 2010; Kojima and Kanoh, 2024). Hereinafter, this paradigm is referred to as the auditory stream segregation, Multiclass, ERP (ASME). Stream segregation is an auditory illusion that makes perceiving alternately presented sounds as segregated multiple sound streams possible (Bregman, 1990; Shamma and Micheyl, 2010; Snyder and Alain, 2007). In the original work (Kanoh et al., 2008, 2010), two different oddball sequences with different pitches were presented simultaneously. The user's target stream was detected by linear discriminant analysis (LDA), which is a two-class system. The sound sequence was presented only to the subject's ear. The detection accuracy reached 95%, and the information transfer rate (ITR) was approximately 5 bits/min. This work was novel compared to that of conventional auditory BCIs in terms of its utilization of human auditory ability. In addition, a multiclass auditory BCI can be created with monaural channel sounds, and it can be used by a user who has an impairment in one ear.

We proposed a three-class ASME paradigm (Kojima and Kanoh, 2024). However, to realize a four-class ASME paradigm, the simplest solution is to increase the number of streams, i.e., to present four streams, each of which is a classic two-stimulus oddball sequence (see Figure 1A). However, in the ASME paradigm, as the number of streams increases, it becomes difficult to perceive the sequence as segregated streams. The spacing in frequency between streams needs to be wider to make the sequence easier to perceive as segregated streams (Bregman, 1990); however, a wider spacing in frequency widens the stimuli bandwidth (see Figure 2). Since the human audible range is limited to $20 - 20,000$Hz (Rosen and Howell, 2011), the maximum number of streams is limited. One method to increase the number of selections without increasing the number of streams is to include multiple target stimuli in a single oddball sequence. Halder et al. (2010) tested the three-stimulus auditory oddball sequence and achieved high accuracy and ITR. The three-stimulus oddball has two different target stimuli. The four-class ASME paradigm can be achieved with two streams utilizing the three-stimulus oddball (see Figure 1B). Thus, we propose two different strategies for realizing the four-class ASME paradigm: (1) presenting four streams, each of which is a two-stimulus oddball (Figure 1A), and (2) presenting two streams, each of which is a three-stimulus oddball (Figure 1B).

Thus, in this study, two different topics were investigated: (1) the feasibility of the four-class ASME BCI with each strategy and (2) comparing two strategies for the four-class ASME paradigm. The ERP responses were evaluated, and the feasibility of the BCI application was tested by offline analysis and simulation. Additionally, the usability of each paradigm was evaluated by the

FIGURE 1
The concept of an ASME with multiple target stimuli in each stream. **(A)** A four-class ASME paradigm with four streams in which each stream has one target stimulus. **(B)** A four-class ASME paradigm with two streams in which each stream has two target stimuli.

FIGURE 2
The influence of spacing in frequency between streams on stream segregation and bandwidth of stimuli. When the spacing between streams is wide, perceiving the sequence as segregated multiple streams is easier; however, the bandwidth of the stimuli becomes broader.

NASA-TLX (Hart and Staveland, 1988; Hart, 2006; Ortega-Gijon and Mezura-Godoy, 2019).

2　Materials and methods

2.1　Stimuli

In this section, two ASME paradigms, (a) four-stream paradigms with two oddball stimuli and (b) two-stream paradigms with three oddball stimuli, are described.

2.1.1　(a) ASME consisting of four streams with a two-stimuli oddball (ASME-4stream)

Figure 3A shows the sequence for the ASME paradigm, which has four streams with two-stimuli oddballs (ASME-4stream paradigm). S_n are standard stimuli, and D_n are deviant stimuli in stream n. Table 1 shows the frequency of each stimulus. The stimulus onset asynchrony (SOA) was set to $0.15s$. In one trial, 600 stimuli were presented, and the duration was approximately 90 s. In each stream, the presentation ratio was $S_n : D_n = 9 : 1$.

2.1.2　(b) ASME consisting of two streams with a three-stimuli oddball (ASME-2stream)

Figure 3B shows the sequence for the ASME paradigm, which has four streams with a two-stimuli oddball (ASME-2stream paradigm). S_n are standard stimuli, and D_n are deviant stimuli. Table 2 shows the frequency of each stimulus. In this paradigm, each stream had two different deviant stimuli: one had a lower frequency, and the other had a higher frequency. The SOA was set to 0.3 s. In one trial, 300 stimuli were presented, and the duration was approximately 90 s. In each stream, the presentation ratio was $S_n : D_L : D_H = 8 : 1 : 1$, where D_L and D_H are the number of deviant stimuli that had lower and higher frequencies, respectively.

2.1.3 Differences in SOA and presentation ratio between the two conditions

In the ASME-4stream and ASME-2stream conditions, the SOA and presentation ratio were different.

The difference in SOA was made to keep the SOA of the stimulus sequence within the attended stream consistent when selective attention was directed to a specific stream while ignoring others. In a sequence with N streams and an overall SOA of t seconds, the within stream SOA_{stm} becomes $N \times t$. In both the ASME-4stream and ASME-2stream conditions, the SOA_{stm}
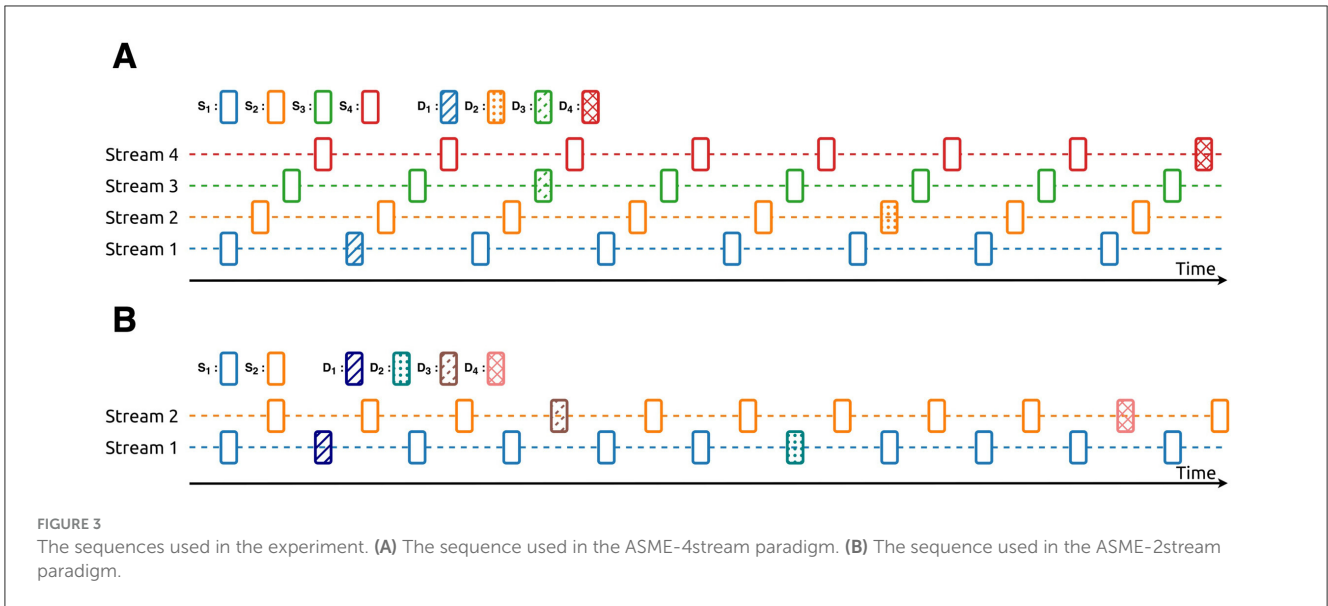
FIGURE 3
The sequences used in the experiment. **(A)** The sequence used in the ASME-4stream paradigm. **(B)** The sequence used in the ASME-2stream paradigm.

TABLE 1 Frequencies for stimuli used for ASME-4stream paradigm.

Stream	Stimulus	Frequency (Hz)
Stream 1	S_1	110.00
	D_1	123.47
Stream 2	S_2	415.30
	D_1	466.16
Stream 3	S_3	1,567.98
	D_3	1,760.00
Stream 4	S_4	5,919.91
	D_4	6,644.86

TABLE 2 Frequencies for the stimuli used for the ASME-2stream paradigm.

Stream	Stimulus	Frequency (Hz)
Stream 1	S_1	523.3
	D_1	440.0
	D_2	622.3
Stream 2	S_2	3,136.0
	D_3	2,637.0
	D_4	3,729.3

was set to 0.6 seconds. Additionally, It has been shown that changes in SOA_{stm} can influence the ease of stream segregation perception (Bregman, 1990), and the SOA_{stm} used in this study was chosen to ensure that each sound stream could be sufficiently perceived as a segregated stream.

The difference in presentation ratio was made to keep the presentation ratio of the target and nontarget stimuli equal when selective attention was paid to a specific deviant stimulus. In the ASME-2stream condition, two deviant stimuli are embedded within each stream. When attention is paid to one of the deviant stimuli, the attended stimulus becomes the target stimulus, and the other stimuli can be treated as nontarget stimuli. For example, in the ASME-2stream condition, the presentation ratio was $S : D_1 : D_2 = 8 : 1 : 1$, however when attention is paid to either D_1 or D_2, the ratio between the target (T) and nontarget stimuli (nT) became $nT : T = 1 : 9$, which was the same as in the ASME-4stream condition.

2.2 Experimental design

Figure 4 shows the structure of an experiment. In the familiarization session, all subjects were presented with both the ASME-4stream and ASME-2stream paradigms. Next, the simple auditory oddball paradigm was conducted with the following parameters in the oddball session. The frequencies of standard and deviant stimuli were 500Hz and 1,000Hz, respectively. The SOA was set to 1.0 s, and the presentation ratio was $S : D = 5 : 1$, where S and D were the number of standard and deviant stimuli, respectively. In the ASME task session, 12 runs of the ASME-2stream and ASME-4stream paradigms were conducted alternately. In a single run, four trials were conducted. Before starting each trial, a screen was placed in front of the subjects and they were instructed on which stream and stimuli to focus on. After the instruction, each standard stimulus of all streams was presented for one second to provide pitch information for the subjects to target the streams better. During the trial, subjects were requested to focus on and count the target deviant stimuli in the target stream.

After the ASME task session, the NASA-TLX questionnaire was administered to score the subjective mental workload for each task. Since all subjects spoke Japanese, the translated NASA-TLX was used (Haga and Mizukami, 1996). With the NASA-TLX, the following six indexes were scored: mental demand, physical demand, temporal demand, performance, effort, and frustration, and the weighted workload (WWL) was obtained.

FIGURE 4
The experimental design.

2.3 Subjects

Fifteen subjects (aged between 21 and 24 years, mean = 22.8 years, two females) were recruited for this study. The study protocol was approved by the Review Board on Bioengineering Research Ethics of the Shibaura Institute of Technology and was conducted in accordance with the Declaration of Helsinki. Before the experiment, the subjects were given information orally and in writing. Written informed consent was obtained from all the subjects. No subject had known cranial nerve diseases or hearing problems.

2.4 EEG and EOG measurement

Sixty-four-channel electroencephalogram (EEG) (Fp1, Fp2, AF7, AF3, AFz, AF4, AF8, F7, F5, F3, F1, Fz, F2, F4, F6, F8, FT9, FT7, FC5, FC3, FC1, FCz, FC2, FC4, FC6, FT8, FT10, T7, C5, C3, C1, Cz, C2, C4, C6, T8, TP9, TP7, CP5, CP3, CP1, CPz, CP2, CP4, CP6, TP8, TP10, P7, P5, P3, P1, Pz, P2, P4, P6, P8, PO7, PO3, POz, PO4, PO8, O1, Oz, and O2) and two-channel (vertical and horizontal) electrooculogram (EOG) were measured by BrainAmp (Brain Products, Germany) at a 1,000 Hz sampling frequency with passive Ag/AgCl electrodes (EasyCap, Germany). The electrodes were placed according to the extended 10-20 system. The reference and ground electrodes were placed on the right and left mastoid, respectively. The subjects sat on a comfortable chair in a soundproofing electromagnetically shielded room.

2.5 EOG artifact removal

EOG artifact removal was conducted as follows. First, the recorded EEG data from each run were highpass filtered by a zero-phase 2nd-order Butterworth filter with a cut-off frequency of 1.0 Hz to remove slow drift and were concatenated along the time domain.

After filtering, principal component analysis (PCA) was applied to the EEG data, and 15 principal components (PCs) were selected.

Next, independent component analysis (ICA) with the FastICA algorithm was applied to the 15 PCs. Two channels of EOG data (vertical and horizontal) were bandpass filtered by a 2nd-order Butterworth filter in the range of 1–10 Hz. The Pearson correlation coefficient between each IC and EOG channel was calculated. The IC, which had the highest correlation with the vertical and horizontal EOG, respectively, was set to zero.

Before applying ICA, 15 components were selected using PCA, as it is known that dimensionality reduction with PCA can improve the quality of the artifact separation (Winkler et al., 2015; Hyvarinen et al., 2001).

2.6 ERP analyses

All measured data after EOG artifact removal were bandpass filtered by a zero-phase 2nd order Butterworth filter in the range of 1–40 Hz. Responses to each stimulus were epoched in the range from −0.1s to 1.2s relative to stimulus onset. Then, all epochs were downsampled to 250Hz. Signed-r^2 values (Blankertz et al., 2011) were obtained to visualize the separability between the responses to the target and nontarget stimuli. The target and nontarget stimuli were defined as follows. All analyses were performed using Python 3.8.18 and MNE-python 1.5.0 (Gramfort et al., 2013).

- ASME-4stream
  The deviant stimuli in the target stream were the target, and the standard stimuli in the target stream, and all stimuli in the nontarget stream were nontarget. For example, when Stream 1 was the target stream, D_1 was the target, and $\{S_1, S_2, S_3, S_4, D_2, D_3, D_4\}$ was the nontarget.
- ASME-2stream
  The attended deviant stimuli in the target stream were the target, and the unattended deviant and standard stimuli in the target stream and all stimuli in the nontarget stream were nontarget. For example, when the target stream was Stream 1 and D_1 was attended, D_1 was the target and $\{S_1, S_2, D_2, D_3, D_4\}$ was the nontarget.

The onset and peak amplitude of P300 responses were estimated with a bootstrap procedure. For this analysis, the following five EEG channels (Fz, FCz, Cz, FPz, and Pz) were used, as it is known that P300 has peak amplitude on these channels (Polich, 2007).

(1) P300 peak amplitude

1. The 80% of the samples of the responses to the target stimuli were taken randomly and averaged across epochs.
2. The peak amplitude in the range of 0.2–0.5 seconds was obtained as v_i.
3. Procedures 1 and 2 were repeated 1,200 times.
4. $\mathbf{v} \in \mathbf{R}^{1,200}$ was averaged and determined as the peak amplitude.

(2) P300 onset latency

1. The 80% of the samples of the responses to the target stimuli (X_{target}) and standard stimuli ($X_{standard}$) were taken randomly.
2. The time sample when the result of the one-sided Welch's t test between X_{target} and $X_{standard}$ was significant ($p < 0.05$) for the

first time and the corresponding time stamp was in between 0.2–0.5 seconds was taken as t_i. If there was no significant difference, it was excluded from the subsequent analysis as a bootstrap sample in which P300 was not observed.

3. Procedures 1 and 2 were repeated 250 times.
4. $\mathbf{t} \in \mathbf{R}^{250-N_f}$ was averaged and determined as the P300 onset latency. N_f is the number of bootstrap samples for which no significant difference was observed in procedure 2.

2.7 Binary classification

After removing EOG artifacts using ICA, all measured data were bandpass filtered by a zero-phase 2nd-order Butterworth filter in the range of 0.1–8 Hz, and responses to each stimulus were epoched in the range of −0.1–1.2 s relative to stimulus onset. Then, all epochs were downsampled to 250 Hz. The mean amplitudes in the ten intervals (0.1 second, non-overlapping intervals from 0 to 1.0 seconds relative to the stimulus onset) were used as the classification feature. The dimensions of the feature vector were 10 intervals × 64 channels = 640. The classification accuracy (AUC: area under the receiver operating characteristic curve) between the responses to the target and nontarget stimuli was obtained by shrinkage linear discriminant analysis (shrinkage-LDA) (Blankertz et al., 2011) with 4-fold chronological cross-validation. For the binary classification, the chance level was 0.5. All analyses were performed with Python 3.8.18, scikit-learn 1.2.0 (Pedregosa et al., 2011). For shrinkage-LDA, implementation included in open source python package toeplitzlda (Sosulski and Tangermann, 2022) 0.2.6 was used.

2.8 BCI simulation (four-class classification)

In the BCI simulation, the target class of each trial out of four classes was estimated. Since six runs were measured for each paradigm, 3-fold chronological cross-validation (two runs for training the machine learning model and four runs for testing) was conducted. First, for training data from two runs, the unmixing and mixing matrix of ICA for removing EOG components was computed by the method described in Section 2.5. Then, the EOG artifact was removed from the training data, the feature vector was obtained using the same method described in Section 2.7, and shrinkage-LDA was trained. The classification output $f(\mathbf{x_i}) = \mathbf{w}^T\mathbf{x_i} + b$ was defined as follows, where $\mathbf{x_i}$ is a feature vector, $\mathbf{w}$ is the weight vector obtained by LDA, and b is a bias. Each feature vector $\mathbf{x_i}$ has a corresponding class label $y_i \in \{-1, 1\}$, and it is assumed that class label +1 is the target and −1 is nontarget. The LDA was trained as $f(\mathbf{x}) \geq 0$ if $\mathbf{x_i}$ was in class +1 and $f(\mathbf{x}) < 0$ if $\mathbf{x_i}$ was in class −1. For the testing data of the four runs, the EOG components were zeroes by the unmixing and mixing matrix of ICA computed with the training data. The feature vectors for each trial were computed from the responses to all deviant stimuli in the trial. Then, the classifier output $f(\mathbf{x_i})$ for each feature was computed, and the class with the largest mean value of classifier output was estimated as the final classification result. The classification results

were evaluated for accuracy. For the BCI simulation, the theoretical chance level was $1/4 = 0.25$. Due to the limited number of samples, the statistical significance of classification accuracy using a binomial cumulative distribution was also evaluated (Combrisson and Jerbi, 2015). Since we had 24 samples' for each condition, it was a 4-class classification, and the threshold for statistical significance of the accuracy P_{th} was $P_{th} = 0.42$ at $p < 0.05$. The detailed method and equations can be found in Combrisson and Jerbi (2015).

All analyses were performed with Python 3.8.18, scikit-learn 1.2.0 (Pedregosa et al., 2011). For shrinkage-LDA, implementation included in open source python package toeplitzlda (Sosulski and Tangermann, 2022) 0.2.6 was used.

The ITR is the amount of information communicated by a system per unit time (Rao, 2019). The ITR can be expressed as follows (Schreuder et al., 2010):

$$R = \log_2(N) + P\log_2(P) + (1-P)\log_2\left(\frac{1-P}{N-1}\right) \qquad (1)$$
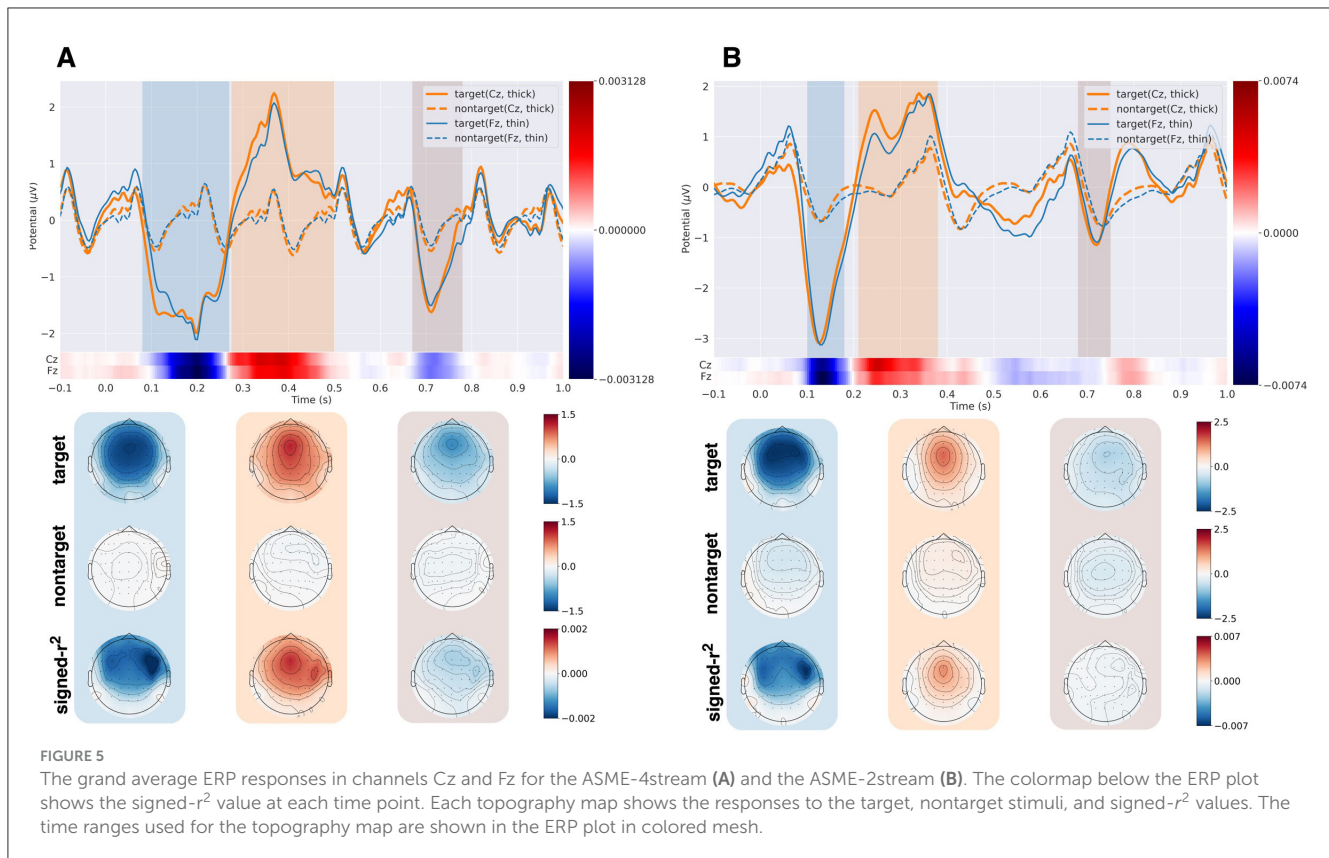
$$B = VR \qquad (2)$$

where N is the number of classes, P is the classification accuracy, V is the classification speed in trials/minute, R is the ITR in bits/trial, and B is the ITR in bits/minute.

3 Results

3.1 ERP analysis

Figure 5 shows the grand average ERP responses to target and nontarget stimuli and their topographic maps for the ASME-4stream and ASME-2stream paradigms, respectively. The signed-r^2 value (Blankertz et al., 2011) between the responses to target and nontarget stimuli was also shown at the bottom of the ERP time courses. For both paradigms, N200 and P300 responses were elicited by the target stimuli. The peak amplitude of the N200 component was greater in the ASME-2stream paradigm, and the peak amplitude of the P300 component was greater in the ASME-4stream paradigm. In both paradigms, N700 responses were also elicited to the target stimuli, where the difference between target and nontarget stimuli is greater in the ASME-4stream paradigm. The N200 responses were elicited front-central in both paradigms. However, the discriminability was greater in the right hemisphere.

Figure 6A shows the P300 peak amplitude of three paradigms, i.e., the ASME-2stream, ASME-4stream, and oddball paradigms, obtained using the bootstrap procedure. The number of bootstrap samples for which no significant difference was observed in the analysis described in the Section 2.6 can be found in Supplementary Table S1. The average peak amplitudes from all subjects were $9.80 \mu V$ (oddball), $3.47 \mu V$ (ASME-2stream), and $3.17 \mu V$ (ASME-4stream). There was no significant difference between the ASME-2stream and ASME-4stream groups ($p = 0.25$, two-sided Wilcoxon signed-rank test). Figure 6B shows the P300 onset latency in the three paradigms. The average latencies from all subjects were 0.23 s (oddball), 0.30 s (ASME-4stream), and 0.24 s (ASME-2stream), and the latencies were significantly greater in

FIGURE 5
The grand average ERP responses in channels Cz and Fz for the ASME-4stream **(A)** and the ASME-2stream **(B)**. The colormap below the ERP plot shows the signed-r^2 value at each time point. Each topography map shows the responses to the target, nontarget stimuli, and signed-r^2 values. The time ranges used for the topography map are shown in the ERP plot in colored mesh.

the ASME-2stream paradigm than in the ASME-4stream paradigm ($p = 0.61 \times 10^{-4}$, two-sided Wilcoxon signed-rank test).

3.2 Classification

The average binary classification scores for all subjects were 0.87 (ASME-2stream) and 0.82 (ASME-4stream). The score in the ASME-2stream was significantly greater ($p = 0.67 \times 10^{-3}$, two-sided Wilcoxon signed-rank test). The highest scores were 0.96 (ASME-2stream) and 0.94 (ASME-4 stream).

Figure 7 shows the accuracy scores and ITR from the BCI simulation results. The average accuracies were 0.86 (ASME-2stream) and 0.83 (ASME-4stream). However, this difference was not significant ($p = 0.40$, two-sided Wilcoxon signed-rank test). The accuracies for both ASME-2stream and ASME-4stream were higher than statistically significant classification accuracy (0.42).

The maximum accuracies were 0.96 (ASME-2stream) and 1.0 (ASME-4stream). The ITRs were 0.87 (ASME-2stream) and 0.82 (ASME-4 stream), which were not significant ($p = 0.49$, two-sided Wilcoxon signed-rank test). The maximum ITRs were 1.16 (ASME-2stream) and 1.32 (ASME-4stream).

3.3 Workload

Figure 8 shows the weighted workload (WWL) obtained by the NASA-TLX questionnaire. The average WLLs of all the subjects

were 61.6 (ASME-2stream) and 71.4 (ASME-4 stram), and these differences were significant ($p = 0.015$, two-sided Wilcoxon signed-rank test). This means that the users' subjective workload was significantly greater in the ASME-4stream paradigm than in the ASME-2stream paradigm. Moreover, negative correlations existed between the classification scores and the *performance* index of NASA-TLX (see Table 3). Note that the *performance* index of the NASA-TLX is on the axis, in which a higher value indicates a lower subjective rating of performance. Among the fifteen subjects, the WWL was lower in the ASME-4stream for two subjects. For these two users, the accuracy of the ASME-4stream BCI simulation was also greater than that of the ASME-2stream BCI simulation.

4 Discussion

4.1 Feasibility of the four-class ASME paradigm

The results of the ASME-4stream paradigm show that it is possible to perceive the presented sequence as four segregated streams and to focus on the target stimuli in the target stream. In addition, the results of the ASME-2stream paradigm show that it is possible to focus on the segregated stream out of two streams and to focus on the target stimulus out of two deviant stimuli. In both paradigms, the accuracy of the BCI simulation (over 0.8 against theoretical chance level = 0.25 and threshold for statistically significant classification accuracy $P_{th} = 0.42$) was high enough for the BCI applications. The ITR in both paradigms was less than
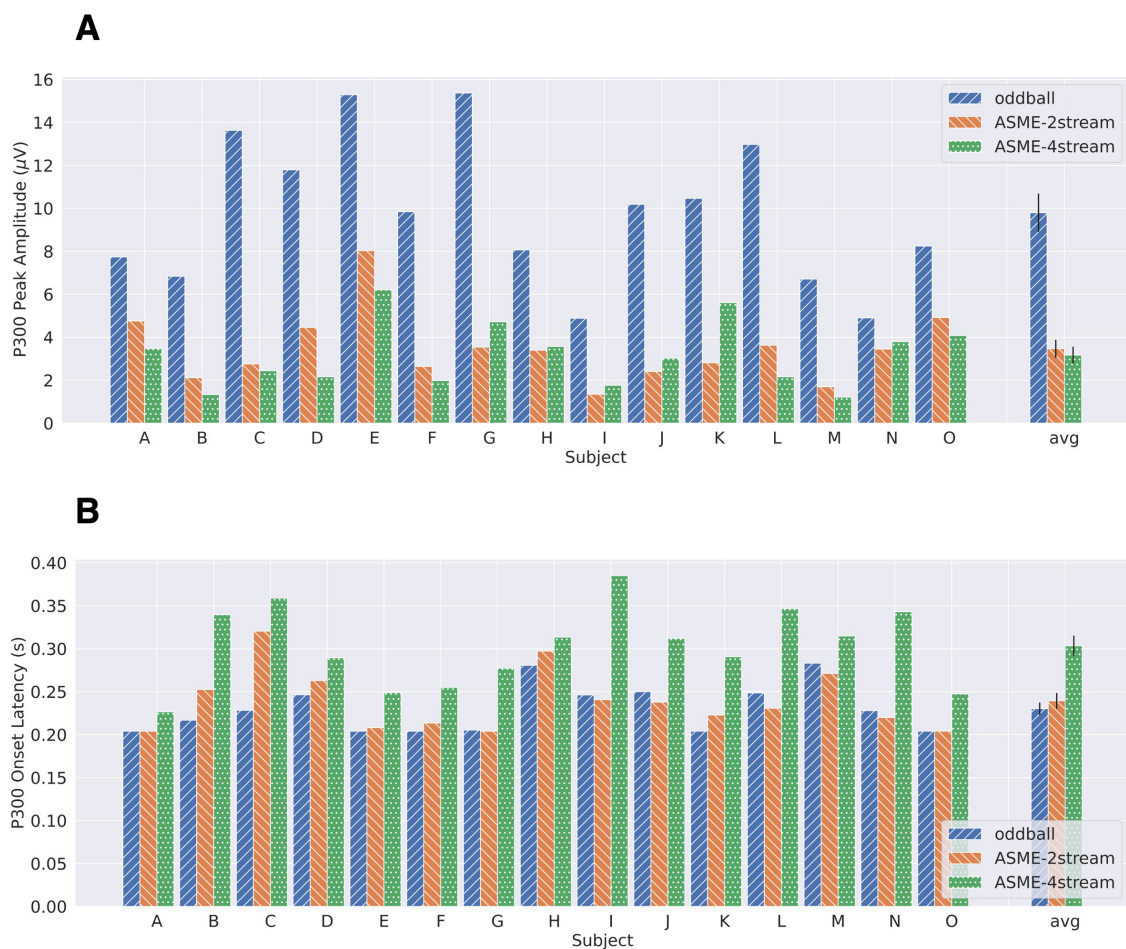
FIGURE 6
The P300 peak amplitudes **(A)** and onset latencies **(B)** for each subject obtained by the bootstrap procedure. Each error bar shown in an averaged bar plot is the standard error mean.
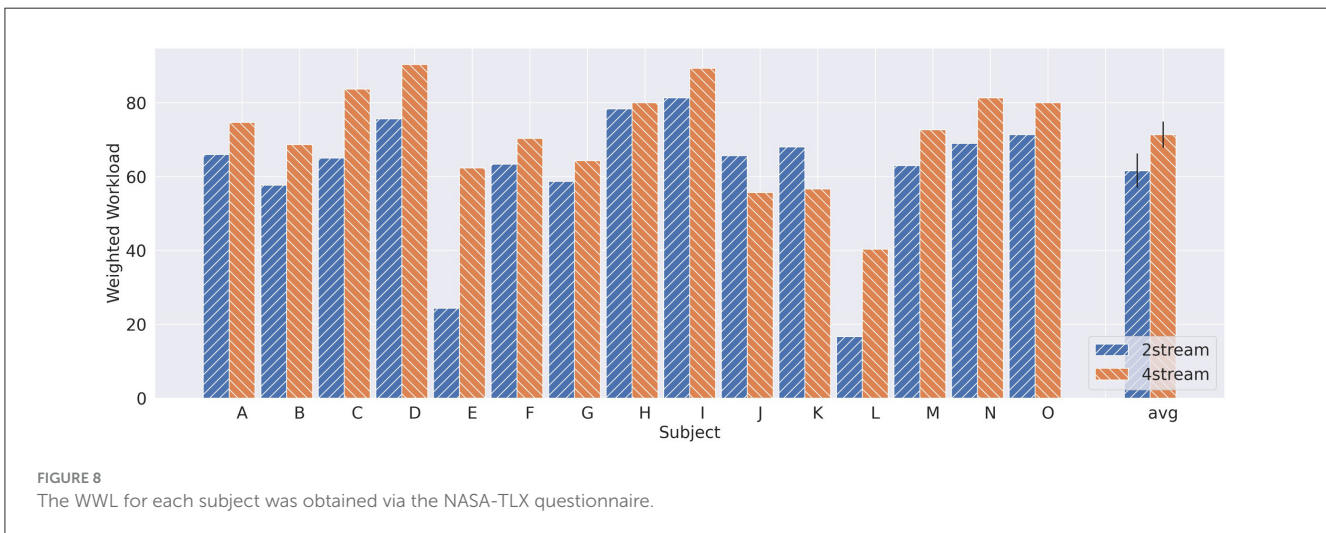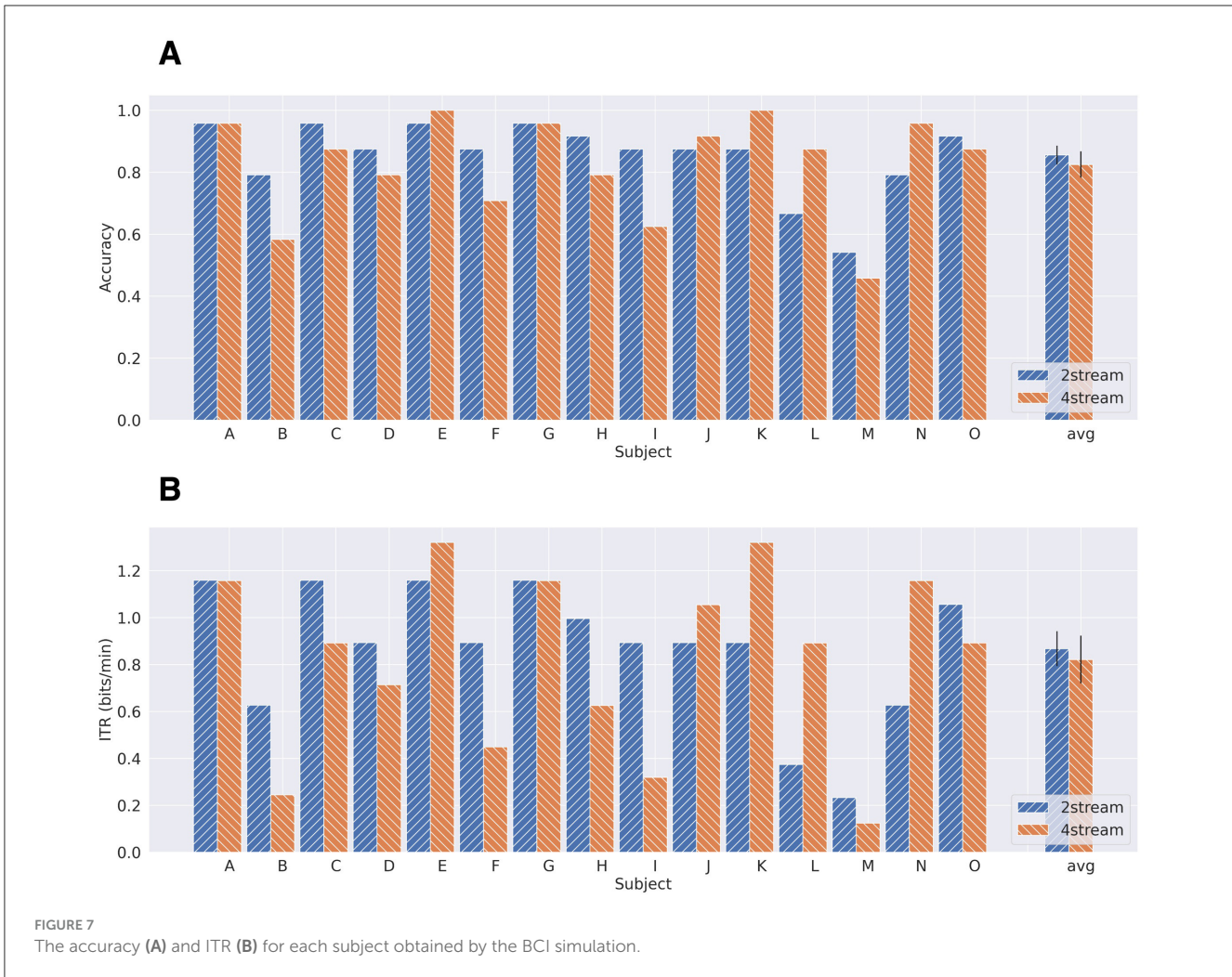
1.0 bits/min and lower than that in other proposed methods (Hill et al., 2004; Furdea et al., 2009; Schreuder et al., 2010; Höhne et al., 2011); however, the trial length was long (90 s) in this study, and it can be further improved by optimizing the trial length by applying dynamic stopping methods (Verschore et al., 2012; Schreuder et al., 2013) or other sophisticated methods. It can be concluded that a four-class BCI system can be used; however, only offline analysis was conducted in this study. Hence, online implementation and evaluation are needed for future work.

4.2 Comparison of two different four-class ASME approaches

ERP analysis with the bootstrapping procedure showed that the average P300 peak amplitude and its onset latency were greater and shorter, respectively, in the ASME-2stream paradigm. Figure 5 shows that the P300 peak amplitude readout was greater in the ASME-4stream paradigm compared to that in the ASME-2stream paradigm. However, the results from P300 peak amplitude estimation obtained by the bootstrap procedure (Section 2.6)

showed larger peak amplitude in ASME-2stream ($3.47\,\mu V$ vs. $3.17\,\mu V$, see Section 3.1). The grand average ERP response waveform in Figure 5 simply shows the average waveform from each subject per EEG channels Cz and Fz. In contrast, with the bootstrap procedure, the peak amplitude of the spatiotemporal feature can be assessed, and it may more accurately reflect the characteristics of ERP peaks. Thus, it can be assumed that the P300 peak amplitude was greater in the ASME-2stream paradigm. Comparing the amplitude and onset latency from the ASME paradigm with those from the simple auditory oddball paradigm, the P300 amplitudes in the oddball paradigm were much larger than those in the two ASME paradigms. However, the P300 onset latency in the ASME-2stream paradigm was on the same level as that in the oddball paradigm, where the P300 latency in the ASME-4stream was significantly greater. According to previous studies (Ghani et al., 2020), the P300 latency was consistently prolonged as task difficulty increased; however, the P300 amplitude was not consistent. Some reported that the P300 increased with increased task difficulty (Combs and Polich, 2006; Muller-Gass and Schröger, 2007); others reported the opposite (Dyke et al., 2015; Horat et al., 2016; Causse et al., 2015; Frank et al., 2012). Hillyard et al. hypothesized that P300 was maximized when "resolution of

FIGURE 7
The accuracy **(A)** and ITR **(B)** for each subject obtained by the BCI simulation.

FIGURE 8
The WWL for each subject was obtained via the NASA-TLX questionnaire.

uncertainty" or "delivery of information" was maximized (Hillyard et al., 1971), which means that tasks that were too clear or too challenging decreased P300 amplitudes; however, the optimal task difficulty maximized the P300 amplitudes. Aggregating these results and findings, it can be concluded that the task difficulty of the ASME-2stream paradigm is lower and more appropriate than that of the ASME-4stream paradigm for eliciting larger P300 amplitudes.

TABLE 3 The correlations between the NASA-TLX performance index and classification performance.

Paradigm	Classification	Correlation	p-value
ASME-4stream	BCI simulation	−0.52	0.046
	Binary classification	−0.48	0.073
ASME-2stream	BCI simulation	−0.64	0.010
	Binary classification	−0.54	0.039

The peak N200 responses were greater in the ASME-2stream paradigm, as shown in Figure 5. However, the average response was broad in the ASME-4stream, suggesting that the latency of the N200 varied across subjects and trials. It was implied that the N200 contributed to the classification due to its large signed-r^2 value.

Frontal-central dominant N700 was also observed in both paradigms; however, the difference between target and nontarget individuals was detected only in the ASME-4stream paradigm. Bender et al. (2010) reported that the amplitude of N700 was enhanced by active short-term memory maintenance compared to attention to current perceptions or passive stimulation. This indicates that more short-term memory maintenance tasks are required in the ASME-4stream paradigm than in the ASME-2stream paradigm. In the ASME-4stream paradigm, the signed-r^2 value for N700 was large, and this component may have contributed to the classification.

The average binary classification and BCI simulation results were better in the ASME-2stream paradigm; however, statistical difference was only observed in the binary classification results. Although the performance is considered comparable for both paradigms, the highest accuracy in this study (accuracy = 1.0 for BCI simulation) was achieved in the ASME-4stream for two subjects, and it was indicated that the ASME-4stream is more suitable for some users to achieve the best performance.

The NASA-TLX results revealed that the subjects' subjective workload was significantly greater in the ASME-4stream paradigm. This suggests that increasing the number of streams increases the user's workload. The classification score and subject's subjective rating of *performance* were also correlated. These results showed that subjective ratings from NASA-TLX can be reflected in quantitative electrophysiological signals and prove the validity of its use in measuring the workload in BCIs. Additionally, this index could indicate which paradigm is the best for the user.

In summary, it can be concluded that the performance is at the same level for both paradigms; however, the user's workload is lower in the ASME-2stream paradigm. Additionally, considering that the best performance was achieved in the ASME-4stream paradigm, using the ASME-4stream paradigm for some users may be adequate to achieve the best performance.

4.3 Conclusion

In this study, it was shown that both paradigms, ASME-2stream and ASME-4stream, can be used as BCI systems with high accuracy. From the results of the ASME-4stream, it was shown that focusing on a single stream out of four segregated streams can be

possible. The results of the ASME-2stream showed that the ASME paradigm involving multiple deviant stimuli in a single stream may be possible. The average performance across subjects was slightly better in the ASME-2stream paradigm (not significant). According to the WWL of the NASA-TLX, the user's workload is lower in the ASME-2stream, and usability is superior in this paradigm. It was shown that the subjects could carry out the task confidently, and the task difficulty was optimal in the ASME-2stream paradigm. However, it was also suggested that determining which paradigm is the best for the subject is encouraged since some subjects achieved greater performance in the ASME-4stream paradigm. It was also shown that sequences with multiple target stimuli in a single stream can be extended to multiple classes with appropriate task difficulty compared to sequences with a single target stimulus in a single stream in the ASME paradigm. These findings expand the possibility of a multiclass extension of the ASME BCI, providing users with choices of practical auditory BCIs.

Data availability statement

All relevant data are publicly available from the Harvard Dataverse repository (https://doi.org/10.7910/DVN/1UJDV6).

Ethics statement

The studies involving humans were approved by Review Board on Bioengineering Research Ethics of the Shibaura Institute of Technology. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

SKo: Conceptualization, Software, Data curation, Formal analysis, Investigation, Visualization, Methodology, Writing – original draft. SKa: Conceptualization, Funding acquisition, Methodology, Project administration, Writing – review & editing, Supervision.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of

their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2024.1461960/full#supplementary-material

References

Belitski, A., Farquhar, J., and Desain, P. (2011). P300 audio-visual speller. *J. Neural Eng.* 8:025022. doi: 10.1088/1741-2560/8/2/025022

Bender, S., Behringer, S., Freitag, C. M., Resch, F., and Weisbrod, M. (2010). Transmodal comparison of auditory, motor, and visual post-processing with and without intentional short-term memory maintenance. *Clin. Neurophysiol.* 121, 2044–2064. doi: 10.1016/j.clinph.2010.05.008

Bin, G., Gao, X., Yan, Z., Hong, B., and Gao, S. (2009). An online multi-channel SSVEP-based brain-computer interface using a canonical correlation analysis method. *J. Neural Eng.* 6:046002. doi: 10.1088/1741-2560/6/4/046002

Blankertz, B., Lemm, S., Treder, M., Haufe, S., and Müller, K.-R. (2011). Single-trial analysis and classification of ERP components — A tutorial. *Neuroimage* 56, 814–825. doi: 10.1016/j.neuroimage.2010.06.048

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound. A Bradford Book*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/1486.001.0001

Causse, M., Fabre, E., Giraudet, L., Gonzalez, M., and Peysakhovich, V. (2015). EEG/ERP as a measure of mental workload in a simple piloting task. *Proc. Manuf.* 3, 5230–5236. doi: 10.1016/j.promfg.2015.07.594

Cheng, M., Gao, X., Gao, S., and Xu, D. (2002). Design and implementation of a brain-computer interface with high transfer rates. *IEEE Trans. Biomed. Eng.* 49, 1181–1186. doi: 10.1109/TBME.2002.803536

Choi, Y.-J., Kwon, O.-S., and Kim, S.-P. (2023). Design of auditory P300-based brain-computer interfaces with a single auditory channel and no visual support. *Cogn. Neurodyn.* 17, 1401–1416. doi: 10.1007/s11571-022-09901-3

Combrisson, E., and Jerbi, K. (2015). Exceeding chance level by chance: the caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *J. Neurosci. Methods* 250, 126–136. doi: 10.1016/j.jneumeth.2015.01.010

Combs, L. A., and Polich, J. (2006). P3a from auditory white noise stimuli. *Clin. Neurophysiol.* 117, 1106–1112. doi: 10.1016/j.clinph.2006.01.023

Dyke, F. B., Leiker, A. M., Grand, K. F., Godwin, M. M., Thompson, A. G., Rietschel, J. C., et al. (2015). The efficacy of auditory probes in indexing cognitive workload is dependent on stimulus complexity. *Int. J. Psychophysiol.* 95, 56–62. doi: 10.1016/j.ijpsycho.2014.12.008

Frank, D. W., Yee, R. B., and Polich, J. (2012). P3a from white noise. *Int. J. Psychophysiol.* 85, 236–241. doi: 10.1016/j.ijpsycho.2012.04.005

Furdea, A., Halder, S., Krusienski, D., Bross, D., Nijboer, F., Birbaumer, N., et al. (2009). An auditory oddball (P300) spelling system for brain-computer interfaces. *Psychophysiology* 46, 617–625. doi: 10.1111/j.1469-8986.2008.00783.x

Gao, Q., Zhao, X., Yu, X., Song, Y., and Wang, Z. (2018). Controlling of smart home system based on brain-computer interface. *Technol. Health Care* 26, 769–783. doi: 10.3233/THC-181292

Gao, X., Xu, D., Cheng, M., and Gao, S. (2003). A BCI-based environmental controller for the motion-disabled. *IEEE Trans. Neural Syst. Rehabil. Eng.* 11, 137–140. doi: 10.1109/TNSRE.2003.814449

Ghani, U., Signal, N., Niazi, I. K., and Taylor, D. (2020). ERP based measures of cognitive workload: a review. *Neurosci. Biobehav. Rev.* 118, 18–26. doi: 10.1016/j.neubiorev.2020.07.020

Gramfort, A., Luessi, M., Larson, E., Engemann, D., Strohmeier, D., Brodbeck, C., et al. (2013). MEG and EEG data analysis with MNE-Python. *Front. Neurosci.* 7:267. doi: 10.3389/fnins.2013.00267

Guger, C., Grünwald, J., and Xu, R. (2020). "Noninvasive and invasive BCIs and hardware and software components for BCIs," in *Handbook of Neuroengineering*, ed. N. V. Thakor (Singapore: Springer), 1–32. doi: 10.1007/978-981-15-2848-4_34-2

Guger, C., Spataro, R., Allison, B. Z., Heilinger, A., Ortner, R., Cho, W., et al. (2017). Complete locked-in and locked-in patients: command following assessment and communication with vibro-tactile P300 and motor imagery brain-computer interface tools. *Front. Neurosci.* 11:251. doi: 10.3389/fnins.2017.00251

Haga, S., and Mizukami, N. (1996). Japanese version of NASA task load index. *Japanese J. Ergon.* 32, 71–79. doi: 10.5100/jje.32.71

Halder, S., Rea, M., Andreoni, R., Nijboer, F., Hammer, E. M., Kleih, S. C., et al. (2010). An auditory oddball brain-computer interface for binary choices. *Clin. Neurophysiol.* 121, 516–523. doi: 10.1016/j.clinph.2009.11.087

Hart, S. G. (2006). Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proc. Hum. Factors Ergon. Soc. Annual Meeting* 50, 904–908. doi: 10.1177/154193120605000909

Hart, S. G., and Staveland, L. E. (1988). "Development of NASA-TLX (Task Load Index): results of empirical and theoretical research," in *Advances in Psychology*, eds. P. A. Hancock, and N. Meshkati (North-Holland: Human Mental Workload), 139–183. doi: 10.1016/S0166-4115(08)62386-9

Hill, N., Lal, T., Bierig, K., Birbaumer, N., and Schölkopf, B. (2004). "An auditory paradigm for brain-computer interfaces," in *Advances in Neural Information Processing Systems* (MIT Press).

Hillyard, S. A., Squires, K. C., Bauer, J. W., and Lindsay, P. H. (1971). Evoked potential correlates of auditory signal detection. *Science* 172, 1357–1360. doi: 10.1126/science.172.3990.1357

Höhne, J., Schreuder, M., Blankertz, B., and Tangermann, M. (2011). A novel 9-class auditory erp paradigm driving a predictive text entry system. *Front. Neurosci.* 5:99. doi: 10.3389/fnins.2011.00099

Holzner, C., Guger, C., Grönegress, C., Edlinger, G., and Slater, M. (2009). "Using a P300 brain computer interface for smart home control," in *World Congress on Medical Physics and Biomedical Engineering, September 7 - 12, 2009, Munich, Germany, IFMBE Proceedings*, eds. O. Dössel, and W. C. Schlegel (Berlin, Heidelberg: Springer), 174–177. doi: 10.1007/978-3-642-03889-1_47

Horat, S. K., Herrmann, F. R., Favre, G., Terzis, J., Debatisse, D., Merlo, M. C. G., et al. (2016). Assessment of mental workload: a new electrophysiological method based on intra-block averaging of ERP amplitudes. *Neuropsychologia* 82, 11–17. doi: 10.1016/j.neuropsychologia.2015.12.013

Hyvarinen, A., Karhunen, J., and Oja, E. (2001). *Independent Component Analysis*. New York: J. Wiley. doi: 10.1002/0471221317

Kanoh, S., Miyamoto, K.-, i., and Yoshinobu, T. (2008). "A brain-computer interface (BCI) system based on auditory stream segregation," in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 642–645. doi: 10.1109/IEMBS.2008.4649234

Kanoh, S., Miyamoto, K.-I, and Yoshinobu, T. (2010). A brain-computer interface (BCI) system based on auditory stream segregation. *J. Biomechan. Sci. Eng.* 5, 32–40. doi: 10.1299/jbse.5.32

King, C. E., Wang, P. T., Chui, L. A., Do, A. H., and Nenadic, Z. (2013). Operation of a brain-computer interface walking simulator for individuals with spinal cord injury. *J. Neuroeng. Rehabil.* 10:77. doi: 10.1186/1743-0003-10-77

Kojima, S., and Kanoh, S. (2024). An auditory brain-computer interface based on selective attention to multiple tone streams. *PLoS ONE* 19:e0303565. doi: 10.1371/journal.pone.0303565

Kosmyna, N., Tarpin-Bernard, F., Bonnefond, N., and Rivet, B. (2016). Feasibility of BCI control in a realistic smart home environment. *Front. Hum. Neurosci.* 10:416. doi: 10.3389/fnhum.2016.00416

Kübler, A., Furdea, A., Halder, S., Hammer, E. M., Nijboer, F., and Kotchoubey, B. (2009). A brain-computer interface controlled auditory event-related potential (P300) spelling system for locked-in patients. *Ann. New York Acad. Sci.* 1157, 90–100. doi: 10.1111/j.1749-6632.2008.04122.x

Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique*. Cambridge, MA: The MIT Press.

Martínez-Cagigal, V., Thielen, J., Santamaría-Vázquez, E., Pérez-Velasco, S., Desain, P., and Hornero, R. (2021). Brain-computer interfaces based on code-modulated visual evoked potentials (c-VEP): a literature review. *J. Neural Eng.* 18:061002. doi: 10.1088/1741-2552/ac38cf

Muller-Gass, A., and Schröger, E. (2007). Perceptual and cognitive task difficulty has differential effects on auditory distraction. *Brain Res.* 1136, 169–177. doi: 10.1016/j.brainres.2006.12.020

Musso, M., Hübner, D., Schwarzkopf, S., Bernodusson, M., LeVan, P., Weiller, C., et al. (2022). Aphasia recovery by language training using a brain-computer interface: a proof-of-concept study. *Brain Commun*. 4:fcac008. doi: 10.1093/braincomms/fcac008

Oralhan, Z. (2019). A new paradigm for region-based p300 speller in brain computer interface. *IEEE Access* 7, 106618–106627. doi: 10.1109/ACCESS.2019.2933049

Ortega-Gijon, Y. N., and Mezura-Godoy, C. (2019). "Usability evaluation of brain computer interfaces: analysis of methods and tools," in *2019 IEEE International Conference on Engineering Veracruz (ICEV)*, 1–8. doi: 10.1109/ICEV.2019.8920409

Pan, X., Zhang, Z., Qu, J., Zhao, C., and Li, Y. (2017). "Enjoy driving from thought in a virtual city," in *2017 36th Chinese Control Conference (CCC)*, 11034–11040. doi: 10.23919/ChiCC.2017.8029119

Park, S., Cha, H.-S., and Im, C.-H. (2019). Development of an online home appliance control system using augmented reality and an SSVEP-based brain-computer interface. *IEEE Access* 7, 163604–163614. doi: 10.1109/ACCESS.2019.2952613

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in Python. *J. Mach. Lear. Res*. 12, 2825–2830.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clin. Neurophysiol*. 118, 2128–2148. doi: 10.1016/j.clinph.2007.04.019

Rao, R. P. N. (2019). *Brain-Computer Interfacing: An Introduction*. Cambridge New York Melbourne New Delhi Singapore: Cambridge University Press.

Rosen, S., and Howell, P. (2011). *Signals and Systems for Speech and Hearing*. New York: Emerald, Bingley, 2. ed edition.

Schreuder, M., Blankertz, B., and Tangermann, M. (2010). A new auditory multi-class brain-computer interface paradigm: spatial hearing as an informative Cue. *PLoS ONE* 5:e9813. doi: 10.1371/journal.pone.0009813

Schreuder, M., Höhne, J., Blankertz, B., Haufe, S., Dickhaus, T., and Tangermann, M. (2013). Optimizing event-related potential based brain-computer interfaces: a systematic evaluation of dynamic stopping methods. *J. Neural Eng*. 10:036025. doi: 10.1088/1741-2560/10/3/036025

Shamma, S. A., and Micheyl, C. (2010). Behind the scenes of auditory perception. *Curr. Opin. Neurobiol*. 20, 361–366. doi: 10.1016/j.conb.2010.03.009

Snyder, J. S., and Alain, C. (2007). Toward a neurophysiological theory of auditory stream segregation. *Psychol. Bull*. 133, 780–799. doi: 10.1037/0033-2909.133.5.780

Sosulski, J., and Tangermann, M. (2022). Introducing block-Toeplitz covariance matrices to remaster linear discriminant analysis for event-related potential brain-computer interfaces. *J. Neural Eng*. 19:066001. doi: 10.1088/1741-2552/ac9c98

Thielen, J., Marsman, P., Farquhar, J., and Desain, P. (2021). From full calibration to zero training for a code-modulated visual evoked potentials for brain-computer interface. *J. Neural Eng*. 18:056007. doi: 10.1088/1741-2552/abecef

Verschore, H., Kindermans, P.-J., Verstraeten, D., and Schrauwen, B. (2012). "Dynamic stopping improves the speed and accuracy of a P300 speller," in *Artificial Neural Networks and Machine Learning — ICANN 2012, Lecture Notes in Computer Science*, eds. A. E. P. Villa, W. Duch, P. Érdi, F. Masulli, and G. Palm (Berlin, Heidelberg: Springer), 661–668. doi: 10.1007/978-3-642-33269-2_83

Vidal, J. J. (1973). Toward direct brain-computer communication. *Ann. Rev. Biophys. Bioeng*. 2, 157–180. doi: 10.1146/annurev.bb.02.060173.001105

Winkler, I., Debener, S., Müller, K.-R., and Tangermann, M. (2015). "On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 4101–4105. doi: 10.1109/EMBC.2015.7319296

Wolpaw, J. R., Birbaumer, N., Heetderks, W. J., McFarland, D. J., Peckham, P. H., Schalk, G., et al. (2000). Brain-computer interface technology: a review of the first international meeting. *IEEE Trans. Rehab. Eng*. 8, 164–173. doi: 10.1109/TRE.2000.847807

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., and Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clin. Neurophysiol*. 113, 767–791. doi: 10.1016/S1388-2457(02)00057-3

Yu, Y., Zhou, Z., Yin, E., Jiang, J., Tang, J., Liu, Y., et al. (2016). Toward brain-actuated car applications: self-paced control with a motor imagery-based brain-computer interface. *Comput. Biol. Med*. 77, 148–155. doi: 10.1016/j.compbiomed.2016.08.010

Zhang, R., Wang, Q., Li, K., He, S., Qin, S., Feng, Z., et al. (2017). A BCI-based environmental control system for patients with severe spinal cord injuries. *IEEE Trans. Biomed. Eng*. 64, 1959–1971. doi: 10.1109/TBME.2016.2628861